Livrable D2.3 of the PERSEE project - Texture analysis/synthesis: Softwares results and final report.

Josselin Gautier, Olivier Le Meur

Projet PERSEE
SCHÉMAS PERCEPTUELS ET CODAGE VIDÉO 2D ET 3D
n° ANR-09-BLAN-0170

Livrable **D2.3** *01/07/2013*

Texture analysis/synthesis: Softwares results and final report.

Josselin GAUTIER IRISA
Olivier LE MEUR IRISA

Contents

1 Introduction

This document dresses a list of the methods which have been designed over the project PERSEE.
In section II, an image inpainting method is presented. Section III is dedicated to a view synthesis method based on image inpainting.

2 Image inpainting

2.1 Introduction

Image inpainting refers to methods which consist in filling in missing regions (holes) in an image [2]. Existing methods can be classified into two main categories.

The first category concerns diffusion-based approaches which propagate linear structures or level lines (so-called isophotes) via diffusion based on partial differential equations [2, 17] and variational methods [3]. The diffusion-based methods tend to introduce some blur when the hole to be filled in is large. The second family of approaches concerns examplar-based methods which sample and copy best matching texture patches from the known image neighbourhood [5, 9, 13, 1]. These methods have been inspired from texture synthesis techniques [10] and are known to work well in cases of regular or repeatable textures. The first attempt to use examplar-based techniques for object removal has been reported in [13]. The authors in [9] improve the search for similar patches by introducing an a priori rough estimate of the inpainted values using a multi-scale approach which then results in an iterative approximation of the missing regions from coarse-to-fine levels. The two types of methods (diffusion- and examplar-based) can be efficiently combined, e.g. by using structure tensors to compute the priority of the patches to be filled in as in [15]. A recent approach [16] combines an examplar-based approach with super-resolution. It is a two-steps algorithm. First a coarse version of the input picture is inpainted. The second step consists in creating an enhanced resolution picture from the coarse inpainted image.

Although tremendous progress has been made in the past years on examplar-based inpainting, there still exists a number of problems. We believe that the most important one is related to the parameter settings such as the filling order and the patch size.

This problem is here addressed by considering multiple inpainted versions of the input image. To generate this set of inpainted pictures, different settings are used. The inpainted pictures are then combined yielding the final inpainted result. Notice that the inpainting algorithm is preferably applied on a coarse version of the input image;

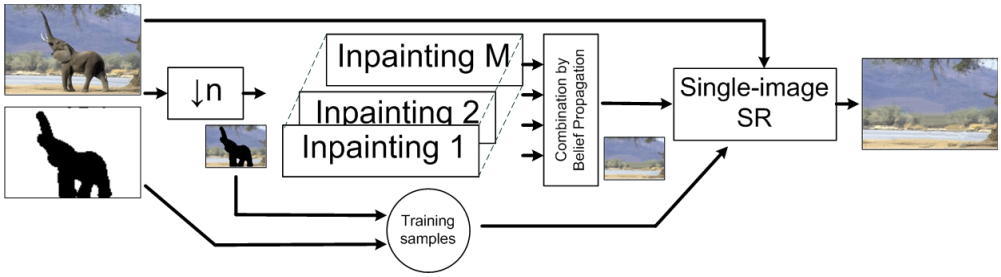


Figure 1: The framework of the proposed method.

this is particularly interesting when the hole to be filled in is large. This provides the advantage to be less demanding in terms of computational resources and less sensitive to noise and local singularities. In this case the final full resolution inpainted image is recovered by using a super-resolution (SR) method similarly to [16]. SR methods refer to the process of creating one enhanced resolution image from one or multiple input low resolution images. These problems are then referred to as single or multiple images SR, respectively. In both cases, the problem is of estimating high frequency details which are missing in the input image(s). The proposed SR-aided inpainting method falls within the context of single-image SR.

The SR problem is ill-posed since multiple high-resolution images can produce the same low-resolution image. Solving the problem hence requires introducing some prior information. The prior information can be an energy functional defined on a class of images which is then used as a regularization term together with interpolation techniques [6]. This prior information can also take the form of example images or of corresponding LR-HR (Low Resolution - High Resolution) pairs of patches learned from a set of un-related training images [11] or from the input low resolution image itself [12]. This latter family of approaches is known as examplar-based SR methods [11]. An examplar-based super-resolution method embedding K nearest neighbours found in an external patch database has also been described in [4]. Instead of constructing the LR-HR pairs of patches from a set of un-related training images, the authors in [12] extract these correspondences by searching for matches across different scales of a multi-resolution pyramid constructed from the input low-resolution image.

The proposed method builds upon the super-resolution-based inpainting method proposed in [16] which is based on examplar-based inpainting (in particular Criminisi-like approach [5]) and single-image examplar-based super-resolution [12]. The main novelty of the proposed algorithm is the combination of multiple inpainted versions of the input picture. The rationale behind this approach is to cope with the sensitivity of examplar-based algorithms to parameters such as the patch size and the filling order. Different combinations have been tested and compared. Besides this major point, different adjustments regarding examplar-based inpainting and SR methods are described

such as the use of the coherence measure to constrain the candidate search [18].

In summary, the proposed method improves on the state-of-the-art examplar-based inpainting methods by proposing a new framework involving a combination of multiple inpainting versions of the input picture followed by a single-image examplar-based SR method. Notice that the SR method is used only when the inpainting method is applied on a low resolution of the input picture.

2.2 Overview

Image completion of large missing regions is a challenging task. As presented in the previous section, there are a number of solutions to tackle the inpainting problem. In this paper, we propose a new inpainting framework relying on both the combination of low-resolution inpainting pictures method and a single-image super-resolution algorithm. In the following sections, we briefly present the main ideas of this paper and the reasons why the proposed method is new and innovative.

The proposed method [14] is composed of two main and sequential operations. The first one is a non-parametric patch sampling method used to fill in missing regions. The inpainting algorithm is preferably applied on a coarse version of the input picture. Indeed a low-resolution picture is mainly represented by its dominant and important structures of the scene. We believe that performing the inpainting of such a low-resolution image is much easier than performing it on the full resolution. A low-resolution image is less contaminated by noise and is composed by the main scene structures. In other words, in this kind of picture, local orientation singularities which could affect the filling order computation are strongly reduced. Second, as the picture to inpaint is smaller than the original one, the computational time is significantly reduced compared to the one necessary to inpaint the full resolution image. To give more robustness, we inpaint the low-resolution picture with different settings (patch's size, filling order, etc). By combining these results, a final low-resolution inpainted picture is obtained. Results will show that the robustness and the visual relevance of inpainting is improved. The second operation is run on the output of the first step. Its goal is to enhance the resolution and the subjective quality of the inpainted areas. Given a low-resolution input image, which is the result of the first inpainting step, we recover its high-resolution using a single-image super-resolution approach. Figure 1 illustrates the main concept underlying the proposed method namely:

1. a low-resolution image is first built from the original picture;

2. an inpainting algorithm is applied to fill in the holes of the low-resolution picture. Different settings are used and inpainted pictures are combined;

3. the quality of the inpainted regions is improved by using a single-image super-

resolution method.

2.3 Software

The executable software is freely available on the following link `http://people.irisa.fr/Olivier.Le_Meur/publi/2013_TIP/`. Note that the archive is password protected. The password will be sent to you on request. The software implements the paper [14].
More information are given on this webpage. In the following section, we just summarize the main inforation.

2.3.1 Usage example

The usage is explained on dedicated webpage. Below we just give an example on the DOS command used to run the software.

```
rem <exe>
rem arg 1: -I <inputFolder>
rem arg 2: -P <picture name without ppm extension>
rem arg 3: -O <outputFolder>
rem arg 4: -x <sizeX>
rem arg 5: -y <sizeY>
rem arg 6: -c <Color the object to remove (BLACK, WHITE or RED)>
rem arg 7: [ OR instead of -c <Color> => -M <picture of the mask without extension pgm>]
rem arg 8: -D <Display On=1 Off=0>
rem arg 9: -H <Number of level>
rem arg 10: -b <save the intermediate low-resolution inpainted pictures yes=1, no=0>

set NB_LEVEL=3
set OUTPUT=C:\LeMeur_TIP2013\results\
set INPUT=C:\LeMeur_TIP2013\input\
set EXE=C:\LeMeur_TIP2013\bin\inpaintingMultipleSR.exe

%EXE% -I %INPUT% -P img1 -O %OUTPUT% -x 512  -y 384   -c BLACK   -D 1 -H %NB_LEVEL%
%EXE% -I %INPUT% -P img2 -O %OUTPUT% -x 384  -y 512    -c BLACK  -D 1 -H %NB_LEVEL%
%EXE% -I %INPUT% -P img3 -O %OUTPUT% -x 720  -y 480   -c BLACK -D 1 -H %NB_LEVEL%
%EXE% -I %INPUT% -P img4 -O %OUTPUT% -x 720  -y 480    -c BLACK  -D 1 -H %NB_LEVEL%
```

3 Inpainting Based View synthesis

3.1 Introduction

3.1.1 View Synthesis

The view synthesis techniques are used at the last stage of the 3D framework before the display of 3D contents. It aims to render new viewpoints thanks to the previously encoded and decoded video plus depth dataset.

A new virtual viewpoint can be generated either by interpolation from multiple views or extrapolation from a single view. Both methods rely on a projection of image plus depth data - or warping - to a new viewpoint: the DIBR.

3.1.2 Template-based Inpainting

Soon after the Bertalmio proposal, based on the work of Efros [10] on texture synthesis, Harrison [13] highlighted the importance of ordering when selecting texture patch(es) from the known neighbourhood instead of single pixels. The exemplar-based or template-based techniques emerged from this idea of copying from the neighbourhood and pasting into the hole some elements of textures or patches according to a certain priority.

Criminisi et al. [5] combined the structure ordering into a textural synthesis: the supposed textured missing regions are inpainted along the isophote direction according to its strength. This method will be described in detail because it constitutes the base of our proposed method.

Criminisi algorithm Criminisi et al. supposed that exemplar-based inpainting contains the essential process to replicate both texture and structure, and confirmed that the structure propagation depends on the order in which the patches are filled-in.

For a given input image $\mathcal{I}$, the source region Φ is defined as the entire image minus the hole region to fill in: $\Phi = \mathcal{I} - \Omega$. A priority $P(p)$ is calculated for all the patches Ψ_p centered at points p along the hole border also called fill front ($p \in \delta\Omega$). This is illustrated in Figure 2. The priority is defined as the product of a confidence term $C(p)$ and a data term $D(p)$:

$$P(p) = C(p)D(p) \quad \text{with} \quad C(p) = \frac{\sum_{q \in \Psi_p \cap (\mathcal{I}-\Omega)} C(q)}{|\Psi_p|}, D(p) = \frac{|\nabla I_p^{\perp} \cdot n_p|}{\alpha} \qquad (1)$$

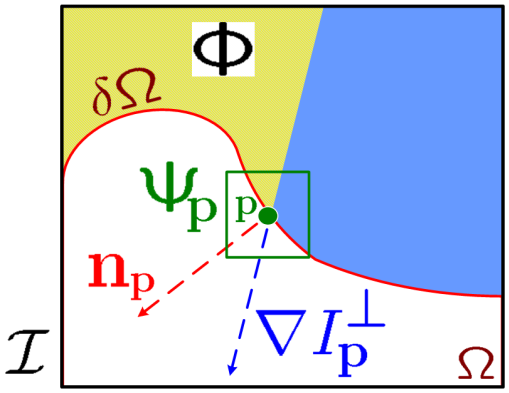


Figure 2: Notation diagram. In the image $\mathcal{I}$, given the patch Ψ_p, n_p is the normal to the contour $\delta\Omega$ of the targeted region Ω and $\nabla I_p^{\perp}$ is the isophote (gradient and intensity) at point p. From [5].

with $|\Psi_p|$ the area of Ψ_p, α a normalization factor, n_p a unit vector orthogonal to the front $\delta\Omega$ in p. $C(p)$ is a measure of the amount of reliable information surrounding the pixel p i.e. the reliability of the current patch. $D(p)$ introduces higher priorities to patch having strong isophotes hitting the front $\delta\Omega$ at each iteration. Then this term boosts the patch having isophote flowing into, thus it favours linear structures to be synthesized and propagated first into the hole region.

Once all the priorities on the fill front have been computed, the patch Ψ_p is selected and the most similar patch $\Psi_{\hat{q}}$ among candidates Ψ_q is derived by:

$$\Psi_{\hat{q}} = \arg \min_{\Psi_q \in \Phi} d(\Psi_{\hat{p}}, \Psi_q) \tag{2}$$

with the distance d defined as the Sum of Square Differences (SSD).

Once the best source examplar $\Psi_{\hat{q}}$ has been found, the value of each remaining pixel into the patch to be filled $p' \in \Psi_{\hat{p}} \cap \Omega$ is copied from its corresponding pixel in $\Psi_{\hat{q}}$ and the confidence term $C(p)$ is updated as follows:

$$C(p) = C(\hat{p}), \quad \forall p \in \Psi_{\hat{p}} \cap \Omega \tag{3}$$

3.1.3 Hybrid methods

Bertalmio et al. proposed in 2003 a new method for simultaneous filling-in of texture and structure (see also [2]), not based on an iterated priority term as in [5] to respect the structure, but instead relies on a separated reconstruction of the structure and texture.

The regions of bounded variation are reconstructed by a diffusion based technique defined in [2]: the propagation by PDE is achieved along the isophote directions. The textural regions are reconstructed separately by a common texture synthesis such as the one proposed in [10].

More recently, Liu et al. proposed a complex but more powerful edge-based inpainting method applied to image compression. Edge extraction and region removal are done accordingly at the encoder side. At decoder side, the exemplar and edge information are first decoded. The non-exemplar regions are classified into structures and textures according to their distances to the edge in the encoder. Structures are generally propagated first by a pixel-wise structure propagation, followed by a common texture synthesis. Indeed, a confidence map similar to that presented above by [5] enables the guidance of the order of structure propagation as well as texture synthesis.

3.1.4 Depth-based View Synthesis by Extrapolation

The presented inpainting technique relies on the search and propagation of the most similar pixel or template based on its colorimetric similarity. In the context of hole-filling for view synthesis from a single view, additional constraints could be added to help to a correct rendering. The past methods and their principles will be first presented before we formalize three statements in order to form a perceptually correct view synthesis. The proposed method will then be detailed.

3.1.5 Past methods

Oh et al. based their method on depth thresholds and boundary region inversion. The foreground boundaries are replaced by the background one located on the opposite side of the hole. Despite the use of two image projections, their algorithm relies on an assumption of connectivity between disoccluded and foreground regions, which may not be verified for high camera baseline configurations. Indeed, over a certain angle and depth, the foreground object does not border the disoccluded part anymore.

Daribo et al. proposed an extension to the Criminisi's [5] algorithm by including the depth in a regularization term for the priority and patch distance calculation. First the patch priority is updated with a level regularity term $L(p)$ defined as the inverse variance of the depth of the current patch Z_p as follows:

$$P(p) = C(p) \cdot D(p) \cdot L(p) \qquad \text{with } L(p) = \frac{|Z_p|}{|Z_p| + \sum_{r \in \Psi_p \cap (\mathcal{I} - \Omega)} (Z_r - \overline{Z_p})^2}$$

with $|Z_p|$ and $\overline{Z}_p$ the area and the mean value of the depth values Z_p respectively. More priority is then given to the patch at the same level, but this doesn't prevent the patch propagation from the foreground. The patch matching was also updated as follows:

$$\Psi_{\hat{q}} = \arg \min_{\Psi_q \in \Phi} d(\Psi_{\hat{p}}, \Psi_q) \; + \; \beta \cdot d(Z_{\hat{p}}, Z_q) \tag{4}$$

with β the factor of importance of depth distance minimization and d the Sum of Square Difference (SSD) distance.

In order to perform the above inpainting operations on the warped view texture containing holes, a knowledge of the corresponding holes Ω in the warped depth is assumed. In practice however, the holes appear both on the texture and depth of the warped view. The Daribo's depth map hole filling proposal consists of using pixel-based Bertalmio [2] inpainting. This approach provides convincing results but still results in visible artefacts .

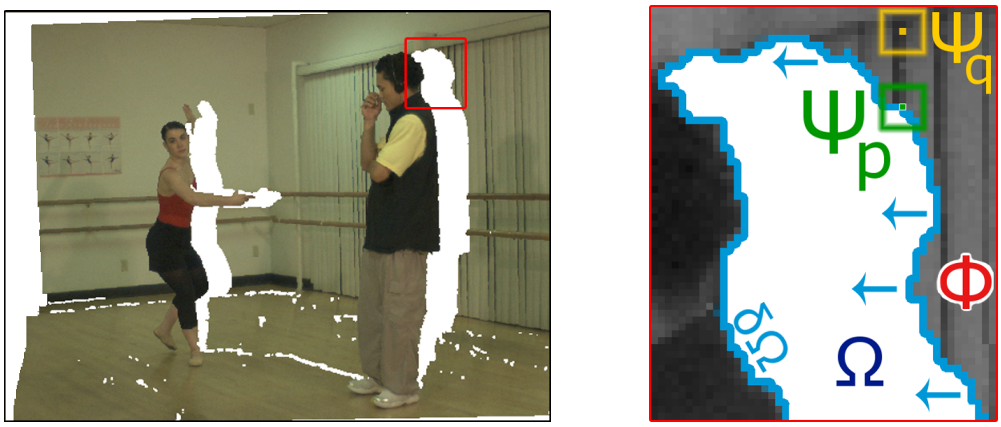
3.2 Overview of the proposed algorithm

The motivation to use a Criminisi-based algorithm resides in its capacity to organize the filling process in a deterministic way. As seen previously and illustrated in Figure 3, this technique propagates similar texture elements $\Psi_{\hat{q}}$ to complete patches Ψ_p along the structure directions, namely the isophotes. This technique then employs the condition of structure and texture propagation defined in last section.

In the context of view synthesis, some of the conditions defined as necessary to a correct perceptual rendering can be directly added as constraints to perform the inpainting. Because the projection in one view will be along the horizontal direction, for a toward-right camera movement the disoccluded parts will appear on the right of their previously occluding foreground (Figure 3a), and oppositely for a toward-left camera movement. Whatever the camera's movement, we have seen that these disoccluded areas should always be filled in with pixels from the background rather than the foreground. Based on this a priori knowledge, we propose a depth-based image completion method for view synthesis based on robust structure propagation.

3.2.1 Tensor-based priority

First, the data term $D(p)$ of the inpainting method proposed by [5] involving the color structure gradient is replaced with a more robust structure tensor. While Criminisi's method favours the patches having isophotes that are "hitting the front" i.e. isophote orthogonal to the orientation of the hole border, we propose to favor the propagation of patches with the strongest isophotes. The approach is different because we aim at



(a) (b)

Figure 3: Illustration of inpainting principle. (a) a warped view, (b) a zoom on the disoccluded area behind the person on the right with the different elements overlaid.

propagating the strongest structure first, whatever its position relative to the border. The tensor term is inspired by partial differential equation (PDE) regularization methods on multivalued images and provides a more coherent local vector orientation. The Di Zenzo matrix [8] is given by:

$$J = \sum_{l=R,G,B} \nabla I_l \nabla I_l^T = \sum_{l=R,G,B} \begin{pmatrix} \frac{\partial I_l}{\partial x}^2 & \frac{\partial I_l}{\partial x}\frac{\partial I_l}{\partial y} \\ \frac{\partial I_l}{\partial x}\frac{\partial I_l}{\partial y} & \frac{\partial I_l}{\partial y}^2 \end{pmatrix}$$

with ∇I_l the local spatial gradient over a 3x3 window. This tensor can also be smoothed with a Gaussian kernel G_σ to be more robust to outliers, without suffering from cancellation effects. We call it $J_\sigma = J * G_\sigma$. Finally, the local vector orientation is computed from the structure tensor J_σ. Its eigenvalues $\lambda_{1,2}$ reflect the amount of structure variation, while its eigenvectors $v_{1,2}$ define an oriented orthogonal basis. Of particular interest is v_2 the preferred local orientation and its "force" λ_2. The reliability of this eigenvector v_2 relative to the gradient to predict structure orientation is illustrated in Figure 4. The eigenvector from a non-smooth tensor is also illustrated. The direction and orientation of v_2 appear to be more reliable than the term $\nabla I_p^\perp$ for calculating the isophote intensity that drives structure propagation. In this example, the structure of the curtain at the background texture will be propagated first because the isophote well matches the direction of the structure to begin propagation. The result leads to a convincing rendering that respects the structure connectivity principles.

More than the eigenvalue λ_2, it is the difference between λ_2 and λ_1 that reflects the

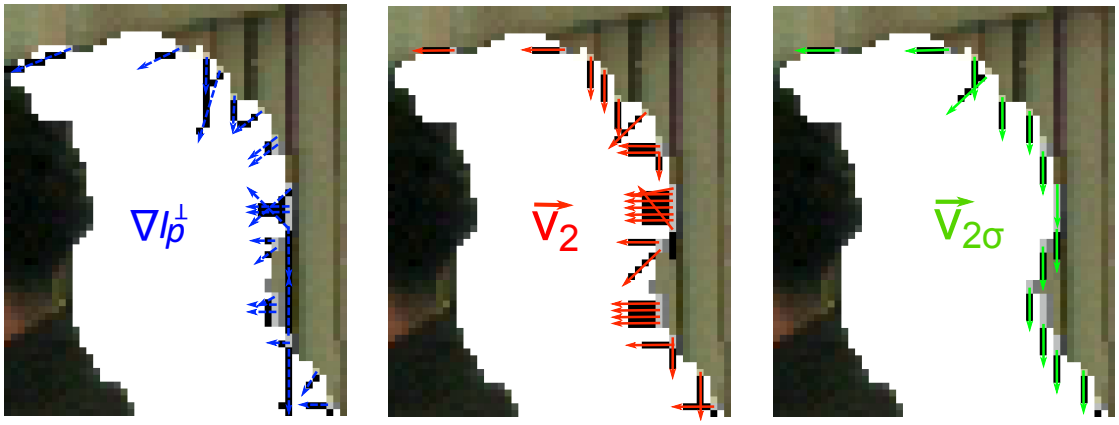


Figure 4: Illustration of the different methods to determine the isophote. For clarity, the black vectors on a pixel grid obtained from the software are overlaid with color vectors. Left: Isophote direction $\nabla I_p^{\perp}$ as computed by [5], as the perpendicular vector to the local gradient. Middle: eigenvector v_2 computed without pre-smoothing of the tensor. Right: v_2 computed with pre-smoothing of the tensor.

intensity of the isophote. Based on the coherence norm, the data term $D(p)$ is then defined as:

$$D(p) = \left[\alpha + (1-\alpha)exp\left(\frac{-C}{(\lambda_1 - \lambda_2)^2}\right)\right] \cdot ||v_2||^2$$

with C a constant positive value and $\alpha \in [0,1]$ (C=8 and α=0.01). Flat regions (when $\lambda_1 \approx \lambda_2$) do not favour any direction, it is isotropic; while with strong edges ($\lambda_1 >> \lambda_2$) the propagation begins along the isophote.

3.2.2 Depth-aided and direction-aided priority

The priority computation has been further improved by exploiting the depth information, first by defining a 3D tensor product, secondly by constraining the side from where to start inpainting.

3D tensor The 3D tensor allows the diffusion of structure not only along color but also along depth information. It is critical to jointly favour color structure as well as geometric structure. The depth-aided structure tensor is extended with the depth map taken as an additional image component Z:

$$J = \sum_{l=R,G,B,Z} \nabla I_l \nabla I_l^T$$

One side only priority The second improvement calculates the traditional priority term along the contour in only one direction. Intuitively, for a camera moving to the right, the disocclusion holes will appear to the right of foreground objects, while out-of-field area will be on the left of the former left border (in white on the left of Figure 3a). We then want to prevent structure propagation from foreground by supporting the directional background propagation, as illustrated in Figure 3b with the blue arrows.

For a projection to the right of the reference view, the patch priority is calculated along the hole border but the top, bottom and left border priorities are set to zero. Globally, the priority term is reset such as:

$$P'(p) = \begin{cases} C(p) \cdot D(p) & \text{if } p \in \delta\Omega_{Dir-projection} \\ 0 & \text{if } p \in \delta\Omega_{Top}, \delta\Omega_{Bottom}, \delta\Omega_{Opp-dir-projection} \end{cases}$$

with $\delta\Omega_{Dir-projection}$, the border pixels on the hole side of direction of projected view, $\delta\Omega_{Opp-dir-projection}$, the opposite side border pixels, $\delta\Omega_{Top}$ and $\delta\Omega_{Bottom}$ the top and bottom side border pixels respectively.

Then for disoccluded areas, the border of the side opposed to direction of view projection is possibly connected to the foreground. It will be filled at the very end of the process. For out-of-field areas, even if left (or right) borders are unknown, we will ensure to begin from the right (or left) border rather than possible top and bottom ones.

These two proposals -extension of the tensor term and non-linear direction inhibition of foreground connected contours- have been included in the prioritization step and are illustrated in Figure 5. For a given iteration, the confidence values $C(p)$ along the hole border appear stronger (darker in 5(b)) in the concavity of the border, because more pixels are known: the confidence is higher. The data term $D(p)$ 5(c) is high everywhere in the neighbourhood of structure, while weak where the local patch is uniform. The resulting multiplication followed by directional inhibition is the priority term $P'(p)$ in 5(d). The top, bottom, and left side of $\delta\Omega$ are effectively reset to a null priority.

3.2.3 Patch matching

Once we precisely know from where to start in a given projected image, it is important to favour the best matching candidates in the background only. Nevertheless, starting from a non-foreground patch does not prevent it from choosing a candidate among the foreground, whatever the distance metric used. Thus, it is crucial to restrict the search to the same depth level in a local window: the background. We simply favour candidates in the same depth range by integrating the depth information in the

(a) Zoom from 3 with $\delta\Omega$ (b) C(p) (c) D(p) (d) P'(p)

Figure 5: Illustration of the different terms used for priority calculation P'(p) along the hole border (in blue in (a)). The intensity of each map in reversed: a darker pixel value means an higher intensity.

commonly used similarity metric L_2-norm Square Sum of Differences (SSD):

$$\Psi_{\hat{q}} = \arg\min_{\Psi_q \in \Phi} d(\Psi_{\hat{p}}, \Psi_q) \quad \text{with } d = \sum_{p,q \in \Psi_{p,q} \cap \Phi} \alpha_l \left\| \Psi_{\hat{p}} - \Psi_q \right\|^2$$

with the same notation as in 2. The depth channel is chosen to be as important as the color one ($l \in R, G, B, Z$ with $\alpha_{R,G,B} = 1$ and $\alpha_Z = 3$). Then it will not prevent the search in foreground patches, but will seriously penalize and unrank the ones having a depth difference greater than i.e in front of the background target patch.

Once the best matching candidates have been found, it is important to combine several of them into a single element to fill in the hole. Directly copying one candidate pixel would lead to a trivial solution that might be not satisfying.

A K-Nearest-Neighbour (KNN) search and combination algorithm is used to compute a robust final patch to fill in iteratively the holes. The patch candidates are not equally reliable and then should not contribute equally to a final combined patch solution. The distance d is translated to a similarity measure s to favour the most similar patch:

$$s(\Psi_{\hat{p}}, \Psi_q) = e^{-\frac{d(\Psi_{\hat{p}}, \Psi_q)}{2\sigma^2}}$$

with σ the factor that will control the smoothness of the patch and d the usual SSD similarity measure. Wexler et al. proposed to choose it as the 75-percentile in the current locations. We calculate it as the squared mean distance of the first and second closest patch, $\sigma = [{}^{(d^1+d^2)}/_2]^2$. then the combined patch Ψ_c is a linear combination of

the K best candidates such as:

$$\forall q \in K \qquad \Psi_c = \frac{\sum_q \Psi_q \; s(\Psi_{\hat{p}}, \Psi_q)}{\sum_q s(\Psi_{\hat{p}}, \Psi_q)}$$

The number of candidates to combine are set in practice to $K = 5$.

Iteration After having filled in a patch, Criminisi et al. [5] updated the priority. For each filling iteration, the priority P(p) was then previously recalculated all along the border of the holes and the patch with the pixel p of highest priority was selected to be filled-in. This method is time-consuming and might lead to favour too much the propagation of the structure.

Our proposal consists in starting the filling from the $N\%$ of pixels along the border with the highest priority, what is called the Percentile Priority-based Concentric Filling (PPCF). Once the patch of the $N\%$ highest priority pixels have been filled in, the priority along the reduced hole border is recalculated. Practically, PPCF allows the maintenance of the propagation of image structure followed by image texture, but to a lesser extent than what is proposed by [5]. The first percent of border pixels of highest priority is used to start the filling in practice.

3.3 Software

3.3.1 Software Instructions and Usage example

The software manual is given in a readme file with the software executable. Below are given the parameters of the DOS command to run the software.

1. INPUT_PATH
2. input_projected_texture-view
3. input_original_depth-map
4. OUTPUT_PATH
5. Resolution_width
6. Resolution_height
7. Level

8. MODE_Weickert
9. MODE_filling
10. 1percent
11. 1_projection_left (1 for projection to the left,0 for projection to the right)

Below are given two examples of inpainting commands for two different view either projected to the right or to the left of the central reference view:

```
5->4 C:\Users\jgautier\Desktop\t\ ballet_color-cam5-4-f000\preproc2\ half
ballet_depth-cam4-f000\half C:\Users\jgautier\Desktop\t\ 512 384 1 W PPCF
1 0

 5->6 C:\Users\jgautier\Desktop\t\ ballet_color-cam5-6-f000\preproc2\ half
ballet_depth-cam6-f000\half C:\Users\jgautier\Desktop\t\ 512 384 1 W PPCF
1 1
```

3.4 Software Results

3.4.1 View Synthesis for Extrapolation: the 3DTV case

In a MVD context with reduced baseline between cameras, we can consider that the extrapolation from a single view located at one or two camera steps from the original camera could provide 3DTV functionality: a subset of views can be rendered within the limited camera range or even one camera step farther. For instance, for a set of acquired views V_1 to V_8, the view in the middle of V_1 and V_2, i.e. $V_{1.5}$ could be easily rendered, but also new extreme virtual views such as V_0 and V_9. Beyond this distance, both the warping and inpainting methods for an extrapolated view synthesis become much more challenging and ambitious.

Figure 6 illustrates the results obtained with the proposed method, compared with methods from the literature [5], [7], when rendering views located at varying distances from the reference viewpoint: one $V_{5\to4}$ at one camera step and one $V_{5\to2}$ at three camera steps from the transmitted reference camera V_5.

The three versions take in input the same warped view texture and the same original view depth information, except for the approach in [5] relying on texture only.

Our method not only preserves the contour of foreground persons, but also successfully reconstructs the structure of missing elements of the disoccluded area (i.e. edges of

the curtains and bars behind the person on the right, background wall behind the left one).

Thanks to our combination term, we can even extend the synthesis to very distant views, without suffering of aliasing effects. As illustrated, the view 5 is projected to view 2 ($V_{5\to 2}$) and the out-of-field blank areas occupying one quarter width of the warped image are reconstructed. The counterpart of the patch combination is the blurring effect appearing on the bottom part of this area. By taking different numbers of patches for combination, it is possible to limit this effect.

(a) $V_{5\to4}$ after warping & background antighosting

(b) $V_{5\to2}$ after warping & background antighosting

(c) $V_{5\to4}$ inpainted with Criminisi's method

(d) $V_{5\to2}$ inpainted with Criminisi's method

(e) $V_{5\to4}$ inpainted with Daribo's method

(f) $V_{5\to2}$ inpainted with Daribo's method

(g) $V_{5\to4}$ inpainted with our method

(h) $V_{5\to2}$ inpainted with our method

Figure 6: Illustration of different methods of inpainting. Our approach relying on 3D tensor and directional prioritization shows efficient filling.

3.5 Importance of the prior depth reconstruction

Following our contribution, Jantet et al. recently proposed improving the Daribo's method by updating the patch matching with the actual full depth knowledge of the patch to fill in with color. The depth map is first projected and the disoccluded pixels are copied only from the background. These depth pixels corresponding to the hole in the texture image are important because they help to find a corresponding similar patch both in texture and in depth in the neighbourhood.

The equation 4 to search for the most similar patch was then updated such as:

$$\Psi_{\hat{q}} = \arg\min_{\Psi_q \in \Phi} d(\Psi_{\hat{p}}, \Psi_q) \; + \; \beta \cdot d_{\mathcal{I}}(Z_{\hat{p}}, Z_q)$$

where the d is the SSD similarity metric performed on the known texture pixel Φ and on the fully known depth pixel of the patch i.e. $\mathcal{I} = \Psi_q \cup \Phi$.

The impact of a correct depth reconstruction and its influence when taking into account in the Daribo's modified method, while not updating our proposal -without full-depth knowledge- is illustrated in Figure 7. This illustrates that an update of the patch matching method significantly improves the original Daribo method and leads to similar results to ours. In perspective it would then be very promising to update the patch matching term in our method with the full depth knowledge per patch.

References

[1] C. Barnes, E. Shechtman, A. Finkelstein, and D. B. Goldman. PatchMatch: A randomized correspondence algorithm for structural image editing. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 28(3), August 2009.

[2] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester. Image inpainting. In *SIGGRPAH 2000*, 2000.

[3] T.F. Chan and J. Shen. Variational restoration of non-flat image features: models and algorithms. *SIAM J. Appl. Math.*, 61(4):1338–1361, 2001.

[4] H. Chang, D-Y Yeung, and Y. Xiong. Super-resolution through neighbor embedding. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, volume I, pages 275–282, 2004.

[5] A. Criminisi, P. Pérez, and K. Toyama. Region filling and object removal by examplar-based image inpainting. *IEEE Trans. On Image Processing*, 13:1200–1212, 2004.

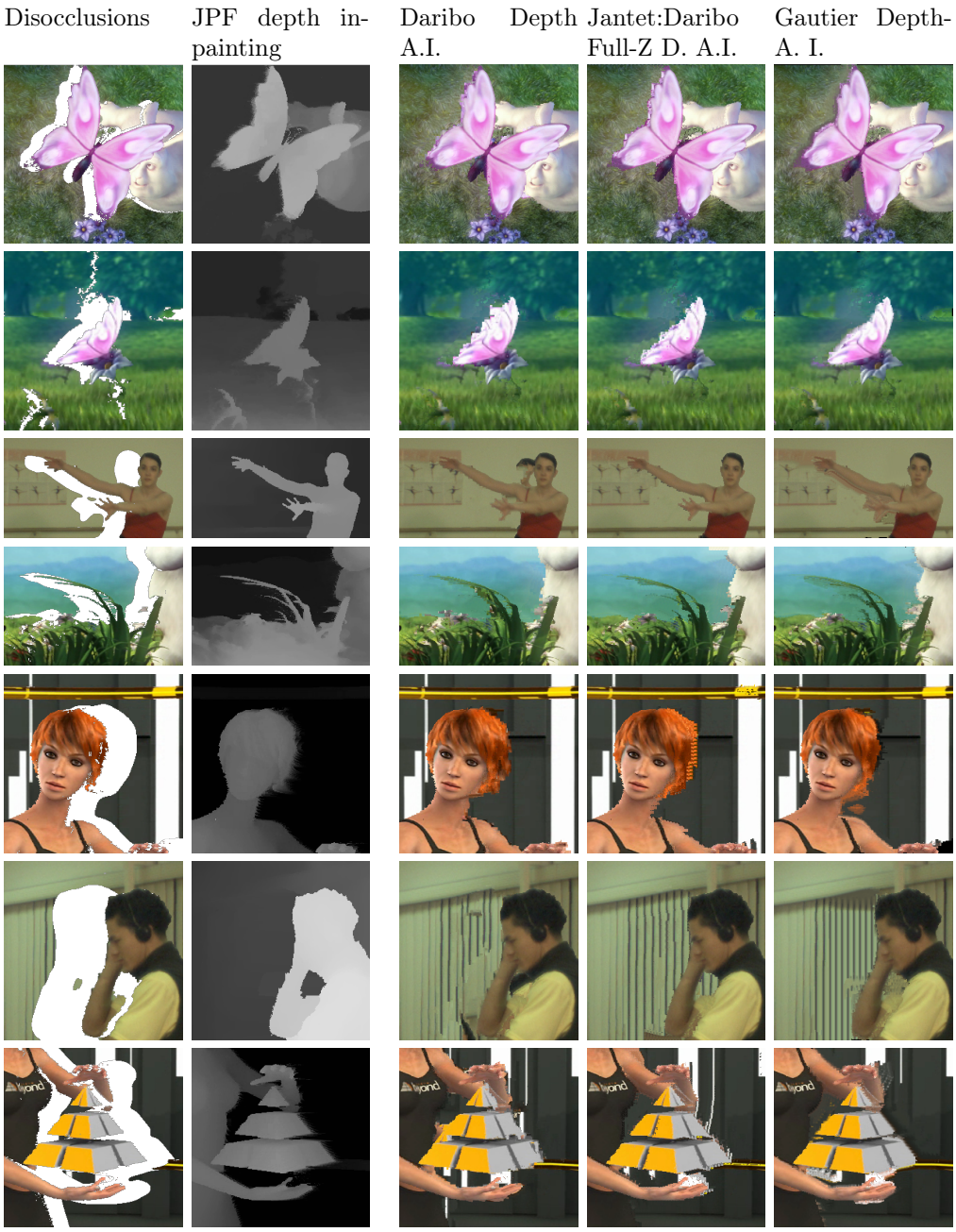


Figure 7: The first column shows a synthesized view with disocclusions. Column 2 illustrates the synthesized depth maps, obtained with the depth inpainting algorithm by backward projection proposed by Jantet et al.. Columns 3, 4 and 5 show the results of three template-based inpaintings based on the original texture shown in column 1 and guided by the depth map presented in column 2.

[6] S. Dai, M. Han, W. Xu, Y. Wu, Y. Gong, and A.K. Katsaggelos. Softcuts: a soft edge smoothness prior for color image super-resolution. *IEEE Trans. On Image Processing*, 18(5):969–981, 2009.

[7] I. Daribo and B. Pesquet-Popescu. Depth-aided image inpainting for novel view synthesis. In *IEEE International Workshop on Multimedia Signal Processing*, 2010.

[8] S. Di Zenzo. A note on the gradient of a multi-image. *Computer Vision, Graphics, and Image Processing*, 33:116–125, 1986.

[9] I. Drori, D. Cohen-Or, and H. Yeshurun. Fragment-based image completion. *ACM Trans. Graph.*, 22(2003):303–312, 2003.

[10] A. A. Efros and T. K. Leung. Texture synthesis by non-parametric sampling. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 1033–1038, 1999.

[11] W. T. Freeman, T. R. Jones, and E. C. Pasztor. Example-based super-resolution. *IEEE Computer Graphics and Applications*, 22(2):56–65, 2002.

[12] D. Glasner, S. Bagon, and M. Irani. Super-resolution from a single image. In *IEEE International Conference on Computer Vision (ICCV)*, volume 10, page 349ñ356, 2009.

[13] P. Harrison. A non-hierarchical procedure for re-synthesis of complex texture. In *Proc. Int. Conf. Central Europe Comp. Graphics, Visua. and Comp. Vision*, 2001.

[14] O. Le Meur, M. Ebdelli, and C. Guillemot. Hierarchical super-resolution-based inpainting. *IEEE TIP*, 2013.

[15] O. Le Meur, J. Gautier, and C. Guillemot. Examplar-based inpainting based on local geometry. In *ICIP*, 2011.

[16] O. Le Meur and C. Guillemot. Super-resolution-based inpainting. In *ECCV*, 2012.

[17] D. Tschumperlé and R. Deriche. Vector-valued image regularization with pdes: a common framework for different applications. *IEEE Trans. on PAMI*, 27(4):506–517, April 2005.

[18] Y. Wexler, E. Shechtman, and M. Irani. Space-time video completion. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, 2004.